



D. Measurement

How well do principals' evaluations of teachers predict student achievement outcomes?

A common element used to measure teacher productivity in performance-based pay systems is student achievement gains, but most performance-pay programs also include the results of teacher evaluations to determine award qualification and amount. A chief concern of teachers is whether principals' evaluations can be objective, accurate, and fair, especially since large portions of performance awards are often determined by these evaluations. Does research suggest that principals' evaluations of teachers are accurate predictors of teacher effectiveness?

A great deal of the early research on principal evaluations examined the potential detrimental effects of basing teacher pay on principals' evaluations of teachers' classroom performance (see review by Milanowski, 2006). The fear was that such assessment was difficult and its potential inaccuracy would limit the motivational impact of performance pay (Murnane & Cohen, 1986). Medley and Coker (1987) did a study of the relationship between evaluation ratings of teachers and their students' achievement, which revealed that the accuracy of principal judgment was low and reinforced this fear. Similarly, Peterson (2000) concluded in a qualitative review of the literature that principals are not accurate evaluators of teacher performance and that both teachers and administrators have little confidence in performance evaluation results.

Research suggests that many principals have a difficult time evaluating teachers, for reasons ranging from lack of knowledge of the subject being taught to disinclination to upset working relationships (Halverson, Kelley, & Kimball, 2004; Nelson & Sassi, 2000; Peterson, 2000). Despite teacher fears that principal evaluations will be unduly harsh, studies suggest that principal evaluations are frequently lenient, and most teachers end up with satisfactory ratings or higher. A recent study of teacher evaluations conducted in Chicago between 2003 and 2006 found that the majority of veteran principals in the district admitted to inflating performance ratings for some of their teachers (The New Teacher Project, 2007). Over the four-year period, 93 percent of Chicago teachers earned the two highest ratings ("superior" or "excellent"), and only 3 in 1,000 received "unsatisfactory" ratings. Even in 87 schools that had been identified as failing, 79 percent did not award a single unsatisfactory rating to teachers between 2003 and 2005.

Although these studies indicate that principals tend to be lenient in practice, other studies suggest that evaluations of teacher performance do predict effectiveness (Murnane, 1975; Armor et al., 1976; Gallagher, 2004; Kimball, White, Milanowski, & Borman, 2004; Milanowski, 2004; Milanowski, Kimball, & Odden, 2005). In one recent study, Jacob and Lefgren (2008) compared principal assessments with measures of teacher effectiveness based on gains in student achievement. The researchers concluded that principals are quite good at identifying teachers whose students make the largest and the smallest standardized achievement gains in their schools

but less able to distinguish between teachers in the middle of the distribution. The difficulty of making finer distinctions between teachers whose performance falls in a broad middle range suggests that states and districts should exercise caution in relying on principals for the finely tuned performance determinations that might be required under certain performance-pay plans. In addition, it should be recognized that the principals in this study did not have to tell the teachers how they were rated and the ratings had no consequences, which may have yielded more accurate and less lenient teacher ratings than might have been observed in a real performance-pay situation. As much prior research in private sector human resources shows, raters are less lenient when the ratings are used for research rather than administrative purposes (see Murphy & Cleveland, 1995).

In summary, the research is somewhat mixed regarding principal accuracy in predicting teacher performance, as measured by the standardized test score results of their students. More recent research suggests that principal evaluations are most accurate at the top and bottom ends of the teacher performance range. However, observations of teachers' classroom performance and standardized test scores measure different dimensions of teacher performance. Principal evaluations can capture important characteristics of effective teaching that test score data cannot, such as a teacher's ability to differentiate instruction. This argues for including principal evaluations as an additional measure of teacher performance rather than basing teacher pay increases solely on student test scores.

Research suggests that principal evaluations have an important role to play in assessing teacher performance. However, it does not tell us how much weight to assign to principal evaluations versus other measures in an overall measure of performance that would be used for teacher compensation. Like many other issues concerning performance pay, states and districts will need to experiment over time with different weights to determine what works best in their particular circumstances. In addition, alleviating teacher concerns about fairness and objectivity will require the use of valid rubrics, or rating scales, to measure desired teacher behaviors; multiple observations of teachers' classroom performance; evaluations conducted by more than one evaluator; and training. Using all of these methods will also help ensure that evaluators' assessments are reliable.

References

- Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976, August). *Analysis of the school preferred reading program in selected Los Angeles minority schools*. R-2007-LAUDS. Santa Monica, CA: RAND Corp. Retrieved January 25, 2008, from <http://www.rand.org/pubs/reports/2005/R2007.pdf>
- Gallagher, H. A. (2004, October). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79(4), 79–107.

- Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In W. Hoy & C. Miskel (Eds.), *Research and theory in educational administration* (Vol. 3). Greenwich, CT: Information Age Publishing.
- Jacob, B. A., & Lefgren, L. (2008, January). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101–136.
- Kimball, S., White, B., Milanowski, A. T., & Borman, G. (2004, July). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education*, 79(4), 54–78.
- Medley, D. M., & Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research*, 80(4), 242–247.
- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33–53.
- Milanowski, A. (2006, October). *Performance pay system preferences of students preparing to be teachers*. WCER Working Paper No. 2006-8. Madison, WI: Consortium for Policy Research in Education, University of Wisconsin. Retrieved January 25, 2008, from http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2006_08.pdf
- Milanowski, A. T., Kimball, S. M., & Odden, A. (2005). Teacher accountability measures and links to learning. In L. Stiefel, A. E. Schwartz, R. Rubenstein, & J. Zabel (Eds.), *Measuring school performance and efficiency: Implications for practice and research* (pp. 137–159). 2005 Yearbook of the American Education Finance Association. Larchmont, NY: Eye on Education, Inc.
- Murnane, R. J. (1975). *The impact of school resources on the learning of inner-city children*. Cambridge, MA: Ballinger.
- Murnane, R., & Cohen, D. (1986). Merit pay and the evaluation problem. *Harvard Educational Review*, 56(1), 1–17.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage Publications, Inc.
- Nelson, B. S., & Sassi, A. (2000). Shifting approaches to supervision: The case of mathematics supervision. *Educational Administration Quarterly*, 36(4), 553–584.
- The New Teacher Project. (2007, July). *Teacher hiring, transfer and assignment in Chicago Public Schools*. New York: Author. Retrieved January 25, 2008, from <http://www.tntp.org/ourresearch/documents/TNTPAnalysis-Chicago.pdf>

Peterson, K. D. (2000) *Teacher evaluation: A comprehensive guide to new directions and practice* (2nd ed.) Thousand Oaks, CA: Corwin Press.

White, B. (2004, April). *The relationship between teacher evaluation scores and student achievement: Evidence from Coventry, RI*. Madison, WI: University of Wisconsin-Madison, Consortium for Policy Research in Education. Retrieved January 25, 2008, from <http://www.wcer.wisc.edu/cpre/papers/CoventryAERA04.pdf>

This synthesis of key research studies was written by:

Cynthia D. Prince, Vanderbilt University; Julia Koppich, Ph.D., J. Koppich and Associates; Tamara Morse Azar, Westat; Monica Bhatt, Learning Point Associates; and Peter J. Witham, Vanderbilt University.

We are grateful to Michael Podgursky, University of Missouri, and Anthony Milanowski, University of Wisconsin-Madison, for their helpful comments and suggestions.